

Amendments to the Specification:

Please replace the paragraph at page 5, line 21 with the following amended paragraph:

In a still further aspect, the present invention provides a process of regulating the function of a nucleotide sequence that contains the sequence 5'-(GNN)n-3' (SEQ ID NO:127), where n is an integer from 1 to 6, the process comprising exposing the nucleotide sequence to an effective amount of a composition of this invention operatively linked to one or more transcription modulating factors. The 5'-(GNN)n-3' sequence can be found in the transcribed region or promotor region of the nucleotide or within an expressed sequence tag. In a preferred embodiment, the nucleotide sequence is part of an oncogene sequence. More preferably, the target nucleotide sequence is contained in a gene that encodes a member of an erbB receptor family. More preferably, the target nucleotide sequence is contained in an erbB gene. Preferred erbB genes are the human erbB-2 and erbB-3 genes.

Please replace the paragraph at page 14, line 20 with the following amended paragraph:

The data show that all possible GNN triplet sequences can be recognized with exquisite specificity by zinc finger domains. Optimized zinc finger domains can discriminate single base differences by greater than 100-fold loss in affinity. While many of the amino acids found in the optimized proteins at the key contact positions -1,3, and 6 are those that are consistent with a simple code of recognition, it has been discovered that optimal specific recognition is sensitive to the context in which these residues are presented. Residues at positions 1,2, and 5 have been found to be critical for specific recognition. Further the data demonstrates for the first time that sequence motifs at positions -1,1, and 2 rather than the simple identity of the position 1

residue are required for highly specific recognition of the 3' base. These residues likely provide the proper stereochemical context for interactions of the helix both in terms of recognition of specific bases and in the exclusion of other bases, the net result being highly specific interactions. Broad utility of these domains would be realized if they were modular in both their interactions with DNA and other zinc finger domains. This could be achieved by working within the likely limitations imposed by target site overlap, namely that sequences of the 5'-(GNN)_n-3' type should be targeted. Ready recombination of the disclosed domains then allows for the creation of polydactyl proteins of defined specificity precluding the need to develop phage display libraries in their generation. These polydactyl proteins have been used to activate and repress transcription driven by the human *erbB-2* promoter in living cells. The family of zinc finger domains described herein is likely sufficient for the construction of 16⁶ or 17 million novel proteins that bind the 5'-(GNN)₆-3' (SEQ ID NO:128) family of DNA sequences.

Please replace the paragraph at page 17, line 1 with the following amended paragraph:

In order to select a family of zinc finger domains recognizing the 5'-GNN-3' subset of sequences, two highly diverse zinc finger libraries were constructed in the phage display vector pComb3H (Barbas III, C. F., Kang, A. S., Lerner, R. A. & Benkovic, S. J. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7978-7982., Rader, C. & Barbas III, C. F. (1997) *Curr. Opin. Biotechnol.* **8**, 503-508). Both libraries involved randomization of residues within the ahelix of finger 2 of C7, a variant of Zif268 (Wu, H., Yang, W.-P. & Barbas III, C. F. (1995) *PNAS* **92**, 344-348). Library 1 was constructed by randomization of positions -1,1,2,3,5,6 using a NNK doping strategy while library 2 was constructed using a VNS doping strategy with randomization of positions -2,-1,1,2,3,5,6. The NNK doping strategy allows for all amino acid combinations within

32 codons while VNS precludes Tyr, Phe, Cys and all stop codons in its 24 codon set. The libraries consisted of 4.4×10^9 and 3.5×10^9 members, respectively, each capable of recognizing sequences of the 5'-GCGNNNGCG-3' type. The size of the NNK library ensured that it could be surveyed with 99% confidence while the VNS library was highly diverse but somewhat incomplete. These libraries are, however, significantly larger than previously reported zinc finger libraries (Choo, Y. & Klug, A. (1994) *Proc Natl Acad Sci U S A* **91**, 11163-7, Greisman, H. A. & Pabo, C. O. (1997) *Science (Washington, D. C.)* **275**, 657-661, Rebar, E. J. & Pabo, C. O. (1994) *Science (Washington, D. C., 1883-)* **263**, 671-3, Jamieson, A. C., Kim, S.-H. & Wells, J. A. (1994) *Biochemistry* **33**, 5689-5695, Jamieson, A. C., Wang, H. & Kim, S.-H. (1996) *PNAS* **93**, 12834-12839, Isalan, M., Klug, A. & Choo, Y. (1998) *Biochemistry* **37**, 12026-33). Seven rounds of selection were performed on the zinc finger displaying-phage with each of the 16 5'-GCGNNNGCG-3' biotinylated hairpin DNAs targets using a solution binding protocol. Stringency was increased in each round by the addition of competitor DNA. Sheared herring sperm DNA was provided for selection against phage that bound non-specifically to DNA. Stringent selective pressure for sequence specificity was obtained by providing DNAs of the 5'-GCGNNNGCG-3' types as specific competitors. Excess DNA of the 5'-GCGNNNGCG-3' type was added to provide even more stringent selection against binding to DNAs with single or double base changes as compared to the biotinylated target. Phage binding to the single biotinylated DNA target sequence were recovered using streptavidin coated beads. In some cases the selection process was repeated. The present data show that these domains are functionally modular and can be recombined with one another to create polydactyl proteins capable of binding 18-bp sequences with subnanomolar affinity. The family of zinc finger domains described herein is sufficient for the construction of 17 million novel proteins that bind the 5'-(GNN)₆-3' (SEQ ID NO:128) family of DNA sequences.

Please replace the paragraph at page 21, line 21 with the following amended paragraph:

The *erbB-2* promoter therefore represents an interesting test case for the development of artificial transcriptional regulators. This promoter has been characterized in detail and has been shown to be relatively complex, containing both a TATA-dependent and a TATA-independent transcriptional initiation site (Ishii, S., Imamoto, F., Yamanashi, Y., Toyoshima, K. & Yamamoto, T. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4374-4378). Whereas early studies showed that polydactyl proteins could act as transcriptional regulators that specifically activate or repress transcription, these proteins bound upstream of an artificial promoter to six tandem repeats of the proteins binding site (Liu, Q., Segal, D. J., Ghiara, J. B. & Barbas III, C. F. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 5525-5530). Furthermore, this study utilized polydactyl proteins that were not modified in their binding specificity. Herein, we tested the efficacy of polydactyl proteins assembled from predefined building blocks to bind a single site in the native *erbB-2* promoter. Described above is the generation and characterization of a family of zinc finger domains that bind each of the 16 5'-(GNN)-3' DNA triplets. One reason we focused on the production of this family of recognition domains is that promoter regions of most organisms are relatively GC rich in their base content. Thus, if proteins recognizing 5'-(GNN)_n-3' sites could be readily assembled from this set of defined zinc finger domains, many genes could be rapidly and specifically targeted for regulation. A protein containing six zinc finger domains and recognizing 18 bp of DNA should be sufficient to define a single address within all known genomes. Examination of the *erbB-2* promoter region revealed two 5'-(GNN)₆-3' (SEQ ID NO:128) sites and one 5'-(GNN)₉-3' (SEQ ID NO:129) site. One of these sites, identified here as e2c, falls within the 5'-untranslated region of the *erbB-2* gene and was chosen as the target site for the generation of a gene-specific transcriptional switch. A BLAST

sequence similarity search of the GenBank data base confirmed that this sequence is unique to *erbB-2*. The position of the e2c target sequence, downstream and in the vicinity of the two major transcription initiation sites, allowed for the examination of repression through inhibition of either transcription initiation or elongation. An interesting feature of the e2c target site is that it is found within a short stretch of sequence that is conserved between human, rat, and mouse *erbB-2* genes (White, M. R.-A. & Hung, M.-C. (1992) *Oncogene* 7, 677-683). Thus, targeting of this site would allow for the study of this strategy in animal models prior to its application to human disease.

Please replace the paragraph at page 22, line 29 with the following amended paragraph:

The general utility of two different strategies for generating three-finger proteins recognizing 9 bp of DNA sequence was investigated. Each strategy was based on the modular nature of the zinc finger domain, and takes advantage of a family of zinc finger domains recognizing triplets of the 5'-GNN-3'. Two three-finger proteins recognizing halfsites (HS) 1 and 2 of the 5'-(GNN)₆-3' (SEQ ID NO:128) *erbB-2* target site e2c were generated in the first strategy by fusing the pre-defined finger 2 (F2) domain variants together using a PCR assembly strategy. To examine the generality of this approach, three additional three-finger proteins recognizing sequences of the 5'-(GNN)₃-3' type, were prepared using the same approach. Purified zinc finger proteins were prepared as fusions with the maltose binding protein (MBP). ELISA analysis revealed that serially connected F2 proteins were able to act in concert to specifically recognize the desired 9-bp DNA target sequences. Each of the 5 proteins shown was able to discriminate between target and non-target 5'-(GNN)₃-3' sequence.

Please replace the paragraph at page 24, line 8 with the following amended paragraph:

As discussed above, the recognition of 9 bp of DNA sequence is not sufficient to specify a unique site within a complex genome. In contrast, a six-finger protein recognizing 18 bp of contiguous DNA sequence could define a single site in the human genome, thus fulfilling an important prerequisite for the generation of a gene-specific transcriptional switch. Six-finger proteins binding the *erbB-2* target sequence e2c were generated from three-finger constructs by simple restriction enzyme digestion and cloning with F2, Zif268, and Sp1C framework template DNAs. ELISA analysis of purified MBP fusion proteins showed that each of the six-finger proteins was able to recognize the specific target sequence, with little cross reactivity to non-target 5'-(GNN)₆-3' (SEQ ID NO:128) sites or a tandem repeat of the Zif268 target site.

Please replace the paragraph at page 25, line 7 with the following amended paragraph:

The zinc finger domain is generally considered to be modular in nature, with each finger recognizing a 3-bp subsite (Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252**, 809-17). This is supported by our ability to recombine zinc finger domains in any desired sequence, yielding polydactyl proteins recognizing extended sequences of the structure 5'-(GNN)_n-3'. However, it should be noted that at least in some cases, zinc finger domains appear to specify overlapping 4 bp sites rather than individual 3 bp sites. In Zif268, residues in addition to those found at helix positions -1, 3, and 6 are involved in contacting DNA (Elrod-Erickson, M., Rould, M. A., Nekludova, L. & Pabo, C. O. (1996) *Structure* **4**, 1171-1180). Specifically, an aspartate in helix position 2 of F2 plays several roles in recognition and makes a variety of contacts. The carboxylate of the aspartate side chain hydrogen bonds with arginine at position -1, stabilizing its interaction with the 3'-guanine of its target site. This aspartate also participates in water-mediated contacts with the guanine's complementary cytosine. In addition, this carboxylate is observed to make a direct contact

to the N4 of the cytosine base on the opposite strand of the 5'-guanine base of the finger 1 binding site. It is this interaction which is the chemical basis for target site overlap. Indeed, when the Zif268 F2 libraries were selected against the four 5'-GCG GNG GCG-3' sequences, both an arginine at position -1 and an aspartate at position 2 were obtained, analogous to the residues in native Zif268. Since the e2c target sequence (5'-GGG GCC GGA GCC GCA GTG-3') (SEQ ID NO: 112) is followed by an A rather than a G, a potential target site overlap problem was anticipated with finger 1 of an e2c-specific six-finger protein. However, in both the Zif- and SplC-framework six-finger proteins, the GTG-specific finger 1 containing an aspartate at position 2 appears to recognize the sequences 5'-GTGA-3' and 5'-GTGG-3' equally well, as indicated by their very similar affinities to target sites e2c-a and e2c-g.

Please replace the paragraph at page 41, line 7 with the following amended paragraph:

For a detailed analysis of its binding properties, the E3 protein was purified as a fusion with the maltose-binding protein. Initially, an ELISA analysis was carried out, revealing specific binding of the E3 protein to its target site, with little or no crossreactivity to various other 5'-(GNN)₆-3' (SEQ ID NO:128) DNA sequences. A similar observation was made with the E2C protein. However, because of the similarity of the DNA sequences recognized, some crossreactivity of the two proteins with each other's target site was detected. To obtain a quantitative measure for the extent of discrimination between target and nontarget sequence, the affinities of the two proteins to each target sequence was determined by electrophoretic mobility-shift assay.